REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
	Final Technical	7/1/99 - 9/30/01

4. TITLE AND SUBTITLE
Shotgun Sequencing of Plasmids from Marine Sediments

Bacteria - Genetic Exploration

5a. CONTRACT NUMBER

5b. GRANT NUMBER
N00014-99-1-0860

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Jonathan Eisen, Ph.D.

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

The Institute for Genomic Research
9712 Medical Center Drive
Rockville, MD 20850

8. PERFORMING ORGANIZATION REPORT NUMBER
01PR06644-00

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
The Institute for Genomic Research
9712 Medical Center Drive
Rockville, MD 20850

10. SPONSOR/MONITOR'S ACRONYM(S)
TIGR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Unclassified Unlimited

13. SUPPLEMENTARY NOTES

NONE

14. ABSTRACT

We have generated DNA libraries from 10+ plasmids (some samples contained more than one plasmid) and have performed enough DNA sequencing reactions on these estimated to yield approximately 5x sequence coverage. Sequence assemblies were generated for each of these and gap closure performed so far on five plasmids. These five have been analyzed using TIGR's automated bioinformatics tools. All the results are available internally and to our collaborators through a web based system.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Jonathan Eisen, Ph.D.
U	U	U	SAR	2	19b. TELEPHONE NUMBER *(include area code)* 301-838-3507

ANNUAL PROGRESS REPORT

PROJECT TITLE: Shotgun Sequencing of Plasmids from Marine Sediment Bacteria - Genetic Exploration

PRINCIPAL INVESTIGATOR: Jonathan A. Eisen (jeisen@tigr.org)

INSTITUTION: The Institute for Genomic Research

AWARD #: N00014-99-1-0860

REPORTING PERIOD: Final

AWARD PERIOD: 7/1/99 – 9/30/01

OBJECTIVE: To use genome sequencing and analysis methods to characterize a set of cryptic plasmids from marine sediment bacteria.

APPROACH: Plasmids are sequenced using the shotgun strategy, in which plasmids are subcloned into a library of 2000-4000 base pair fragments and then these clones are sequenced randomly. Five-fold shotgun coverage (in which each base pair of the plasmid is sequenced on average 5 times) will be achieved for all plasmids. These shotgun sequences are then used to generate assemblies of plasmid molecules using computer programs than compare each sequence to all others. Gaps will be closed in as many plasmids as is possible given the financial constraints of the program, leading to the generation of complete or nearly complete plasmid assemblies. Gene and genome features of these plasmids are then characterized using a variety of bioinformatics tools.

ACCOMPLISHMENTS We have generated DNA libraries from 10+ plasmids (some samples contained more than one plasmid) and have performed enough DNA sequencing reactions on these estimated to yield approximately 5x sequence coverage. Sequence assemblies were generated for each of these and gap closure performed so far on five plasmids. These five have been analyzed using TIGR's automated bioinformatics tools and results are available internally and to our collaborators on the web.

Greater than 5x coverage was achieved for 9 plasmids. One additional plasmid appears to have been much larger than original estimates and current sequence coverage is only 3-4 fold. The sequences for each plasmid were assembled using TIGR assembler (a program designed at TIGR to assemble shotgun genome sequence data). The resulting assemblies were mostly quite robust and many contained very few gaps. The five plasmids with the best assemblies (least ambiguity and fewest gaps) were then sent to the TIGR closure teams to close as many gaps as possible (closing a gap involves determining the sequence of the region in the gap). All five have now been closed, meaning that we now have the

complete sequence of these plasmids. These five plasmids were then submitted to TIGR's automated annotation process, which involves analysis of gene and genome features including the identification of putative open reading frames, database searches to identify homologs and motifs, and assignment of tentative functions and role categories. All the results are available internally and to our collaborators through a web based system. Analysis of these is continuing with the goal of publishing papers on each or all of the plasmids.

SIGNIFICANCE: The sequence information will be useful for understanding the biology of these plasmids as well as marine plasmids in general. For example, the sequence of the plasmid that is similar to the F plasmid is only the fourth plasmid to be characterized in this plasmid family. The F plasmid is very important to many aspects of the biology of the pathogen *E. coli* and to studies of molecular biology of many species. Therefore this new sequence information will be very helpful in understanding the biology of this family of plasmids. We have also discovered a variety of genes on each plasmid that will be helpful in creating phenotypic screens to study the biology of these plasmids in marine sediment environments.

WORK PLAN (next 12 months): The results are being written up for publication.

PUBLICATIONS, AWARDS AND PATENTS (last 12 months):

None as of yet.